1231868

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

September 29, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: *60/497,814*
FILING DATE: *August 26, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/27432*

Certified by

Jon W Dudas

Acting Under Secretary of Commerce
for Intellectual Property
and Acting Director of the U.S.
Patent and Trademark Office

UNITED STATES PATENT AND TRADEMARK OFFICE
DEPARTMENT OF COMMERCE

Please type a plus sign (+) inside this box → +

Express Mail Label: EL995079201US

PTO/SB/16 (8-00)
Approved for use through 10/31/2002. OMB 0651-0032
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

19587 U.S. PTO 60/497814 08/26/03

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Alexandros	Tourapis	West Windsor, New Jersey
Jill MacDonald	Boyce	Manalapan, New Jersey

☐ *Additional inventors are being named on the _____ separately numbered sheets attached hereto*

TITLE OF THE INVENTION (280 characters max)

METHOD AND APPARATUS FOR ADAPTIVELY SELECTING A NUMBER OF REFERENCE PICTURES FOR ENCODERS

CORRESPONDENCE ADDRESS

Direct all correspondence to:

☐ Customer Number [ ] → *Place Customer Number Bar Code Label here*

OR *Type Customer Number here*

☒ Firm or Individual Name	Joseph S. Tripoli - Thomson Licensing Inc.				
Address	PATENT OPERATIONS				
Address	P. O. BOX 5312				
City	PRINCETON	State	NJ	ZIP	08543-5312
Country	USA	Telephone	609-734-6834	Fax	609-734-6888

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification *Number of Pages* 6 ☐ CD(s), Number

☒ Drawing(s) *Number of Sheets* 1 ☐ Other (specify)

☐ Application Data Sheet. See 37 CFR 1.76

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

☐ Applicant claims small entity status. See 37 CFR 1.27.

☐ A check or money order is enclosed to cover the filing fees

☒ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 07-0832

☐ Payment by credit card. Form PTO-2038 is attached.

FILING FEE AMOUNT ($) 160

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____.

Respectfully submitted,
SIGNATURE [signature]

Date 08/26/03

REGISTRATION NO. 41,736
(if appropriate)

TYPED or PRINTED NAME GUY H. ERIKSEN

Docket Number: PU030257

TELEPHONE (609) 734-6807

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

METHOD AND APPARATUS FOR ADAPTIVELY SELECTING A NUMBER OF REFERENCE PICTURES FOR ENCODERS

BACKGROUND OF THE INVENTION

Recent video coding standards and architectures employ multiple reference pictures for motion estimation and compensation in an attempt to improve coding efficiency. On the other hand, use of multiple references can increase considerably the complexity of the encoder, since more pictures need to be examined during the motion estimation process. Furthermore, since the reference index needs to also be included within the bitstream, (e.g - for every macroblock or macroblock partition; blocks of 16x8, 8x16 or 8x8 pixels), this may imply that it is not always certain that multiple references would provide benefits in encoding a particular picture (i.e. - a picture may be biased towards only a single reference). Considering for example that for each macroblock in H.264 it is possible to transmit up to 4 reference indices for Predictive (P) pictures, and 8 for Bi-directionally (B) predictive pictures, the bitrate overhead due to the reference indices could be quite significant. It would be desirable instead to be able to decide the number of references prior to encoding a given picture, considering that if only one reference is used the above mentioned overhead would be eliminated thus possibly improving encoding performance, while at the same time reducing complexity since fewer references would be tested using motion estimation. In H.264 for example, the number of references is controlled through the **num_ref_idx_l*N*_active_minus1** (*N* is equal to 0 for list0 and 1 for list1) parameter that is signaled t the slice level. If this parameter is equal to 0, then for the current slice, no other reference index information needs to be transmitted for that list.

In previous standards and architectures like MPEG-2 and MPEG-4 only a single reference index was used, while for the encoding of motion vectors a special code was also transmitted within the bitstream for every picture named as the f-code parameter that was used for the determination and decoding of the motion vectors. Essentially this parameter was derived during the motion estimation process, and affected the VLC coding of the motion vectors. Previous proposals for automatically adapting the f_code for every picture, depending on its motion parameters and range, could achieve better coding efficiency, compared to keeping the parameter fixed. Nevertheless, newer standards such as H.264, do not support this parameter,

essentially use predefined VLC codes for the encoding of the motion vectors, and thus no such property could be utilized. On the other hand, H.264 considers multiple references which require a similar parameter to f_code to be transmitted, but no equivalent work has been done up to this point.

SUMMARY OF THE INVENTION

These and other drawbacks and disadvantages of the prior art are addressed by an apparatus and method that enables a video encoder to achieve similar or even better quality, with also lower complexity, by adaptively selecting the number of reference pictures that are used during the encoding process. The decision can be based on previously generated information, such as picture correlation, reference picture motion vectors, residuals, etc, while also this decision could be based on a Rate Distortion Optimal method.

DETAILED DESCRIPTION

In accordance with the principles of the present invention, a new method is presented for deciding the number of references that will be used for the encoding of the current picture.

Obviously a relatively simple method for selecting the number of references for a given picture would be, similar to the method performed for f_code in MPEG-2 and MPEG-4, to encode the picture in a first pass using all possible references, and finally in a second pass recode the picture using only the referenced pictures. An additional consideration could be made on whether the number of macroblocks or blocks that reference a given picture satisfy a given condition/threshold. If this condition is not satisfied, this reference is also removed from the reference buffer, and these macroblocks/blocks are then predicted from the remaining references. Although such methods could potentially lead to better encoding performance, they also introduce considerably higher complexity considering that a picture needs to be coded twice. This is especially burdensome in codecs such as H.264, due to their already very high complexity. Nevertheless, in a more brute force approach, it is possible also to try and encode the same picture K times, using from 1 to M references where K is equal to:

$$K \leq \sum_{i=1..M} \frac{M!}{(M-i)!}$$

which basically implies all possible arrangements and combinations (permutations) of references, including reordering. From these, we can select the one that gives the least distortion, or bitrate, or use rate distortion optimization criteria (e.g. through the use of lagrangian multipliers in the form of $J=D+\lambda\times R$).

It is possible to design a much simpler approach for performing the decision of the number of references, without sacrificing much in encoding quality or bitrate. In particular, we have found that high correlation exists in the indices of the references used between adjacent pictures. Such correlation increases further when the two pictures are of high similarity (e.g. their absolute difference is below a given relatively small threshold). For example, if the immediate previously coded picture, at time t-1, references only picture at t-2, has little motion, and is very similar with the current picture at time t (e.g. picture MAD <4) this suggests that it is also very likely that the current reference would be using a single reference. An additional simple comparison (e.g. absolute difference) between the current picture, and the remaining references could also be performed to further enhance this decision. Finally, as an additional rule, the motion vectors and reference indices not only of the closest but also of all other references could be considered for this decision.

A similar concept could also be applied to B pictures. In this case, considering that usually B pictures are contained within a forward (list0) and a backward (list1) reference we may also add an additional condition depending on the motion vectors and reference indices of both references. For example, it is very likely that if all or a very high percentage of the blocks (i.e. 90%) in the backward reference picture, use the first forward picture as reference, then using only a single reference for list0 can be more beneficial considering the bits saved from not having to code the reference indices. In reality, from experimental results, B pictures do not benefit as much as P pictures from multiple references, considering also the high use of skip modes within this picture type, and this option could be completely disabled without having to perform a reference number decision, and without much impact in quality.

A more specific and rather simple strategy for selecting the number of references, in an exemplary embodiment, is follows:

- Compute the Sum of absolute differences (*SAD*) at the block or macroblock level between the current picture and the first reference in the given list (list0 or list1). This difference could be computed using either the original of the reference, or the reconstructed reference. Compute also the mean absolute difference (*MAD*) for the entire picture. If the *MAD* for the entire picture is below a relatively small threshold T_1 then only one reference is used for that list, and **num_ref_idx_l*N*_active_minus1** is set to 0. If also all or a high percentage greater or equal to *R*% (e.g. *R*%=95%) of the macroblocks have *SAD* value below a threshold T_2 then again a single reference is used, and **num_ref_idx_l*N*_active_minus1** is again set to 0. If the reconstructed reference is used for the distortion calculation, considering that this is also affected by the quantization process, T_1 and T_2 should be adjusted/scaled accordingly. The simplest strategy would be to predefine specific weights that depend on the quantization QP parameters and select T_1 and T_2 as $T_1(QP)=a(QP)\times T_1$ and $T_2(QP)=b(QP)\times T_2$ where $a()$ and $b()$ are the predefined weights.
- If the above rule does not apply, but the MAD of the entire picture satisfies a different threshold T_3 ($T_1 \leq MAD < T_3$) or the SAD for all (or a high percentage of) macroblocks satisfies a different threshold T_4 ($T_2 \leq SAD < T_4$) then we also examine the motion vectors and reference indices used by the first reference of each corresponding list. If all (or a high percentage of e.g. K_1%) reference indices used are equal to zero, then again only a single reference is used, and **num_ref_idx_l*N*_active_minus1** is modified accordingly. Special consideration could also be made on the motion vectors, i.e. if all motion vectors are small enough (low motion activity) to enhance this condition. For example, if a large percentage (e.g. $K_2\% \leq K_1\%$) of the blocks in the picture use the zero reference and at the same time have motion vector components MV_x and MV_y lying in the range of $[-mx_1, mx_2]$ and $[my_1, my_2]$ (e.g $mx_1=mx_2=my_1= my_2=1$) respectively, a single reference is used.
- Otherwise, the remaining references are also compared to the current reference, through the calculation of the entire picture or Block/Macroblock SAD values. If the MAD_i for reference i is above a threshold T_5 or all macroblocks have a value of SAD_i larger than a threshold T_6, then this reference is removed from the examined references. Similar to the previous conditions, motion vectors and reference indices from the closest to further reference can be considered and assist

in the decision, by also adapting the values of T_5 and T_6. In particular, if a reference is not used by another reference that is closer to the current picture, then these thresholds are reduced (reduction implies that reference has higher probability that it is removed from the references examined).

- *Since also the distortion values for a reference compared to its own references might have already been pre-computed for that picture's reference number decision, these distortion values could be reused as an additional decision mechanism. In particular, if it is already known that a reference is very similar to the current picture, but has high distortion compared to another reference, then it is very likely that even the current picture will have high distortion versus that reference and the distortion calculation could be completely avoided, and that reference is automatically removed from the references that will be used. On the other hand, considering the distortion (or the residual if available) between these two references after motion compensation, would probably lead to a better decision and performance.*
- Finally, it is also possible using these generated statistics to perform a reordering of the references (references with smaller distortion are placed with higher priority in the list), i.e. in H.264 by signaling the reference picture list reordering elements (section 7.3.3.1 **Error! Reference source not found.**).

Obviously other methods for estimating distortion could also be used, while this method could be combined with weighted prediction strategies.

For B pictures a similar strategy could be followed. On the other hand, as previously discussed, it is possible to also use both lists for deciding whether a reference will be kept or not. In particular if the list1 of a B picture (e.g. P_9 for pictures B_7 and B_8) uses only a single reference which is also the first reference in list0 (P_6) and there is a temporal relationship between these pictures as can be seen in Figure 1, then it is very likely that these B pictures would also be using a single reference for list0. We may again consider the distortion of these references, but also the motion information, and in particular if most blocks in the list1 reference being stationary or not (having zero or close to zero motion). If the list0 reference also uses a single reference entirely or in its majority, this rule could be strengthened further, while also motion vectors, and distortion between each reference could again be considered.

VARIOUS EMBODIMENTS, ADVANTAGES, ETC.:

1. Perform motion estimation and compensation examining all possible combinations and re-orderings of available references and select the one that minimizes a predetermined condition (rate, distortion, or combination thereof)
2. Perform motion estimation and compensation examining all references, and recoding if only a single reference was used
3. Perform motion estimation and compensation using only a single or a subset of the number of references for a list originally specified in the encoder by examining whether certain criteria are met.
4. As described in number 3 where if only a single or fewer than originally specified references are used, the **num_ref_idx_*lN*_active_minus1** is reduced accordingly.
5. As described in numbers 3 and/or 4, where the distortion between current and first reference picture is used for determining whether a single reference is to be used.
6. As described in number 3 where motion information and references, are used for the decision.
7. As described in number 6 where additional references are also examined with regards to distortion and motion information, and removed if they do not satisfy some given criteria.
8. As described in numbers 3 and/or 4, where motion information/distortion from a different list reference could also be used for determining the number of references that will be used for prediction.

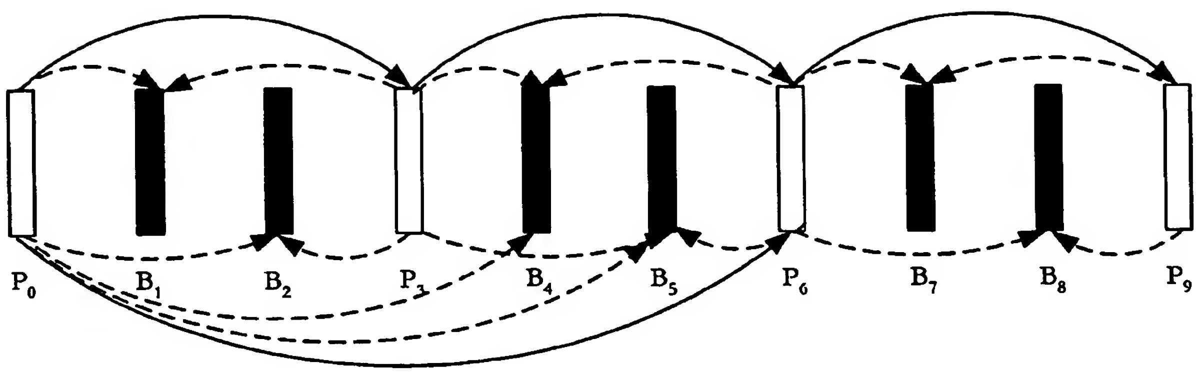

Figure 1: Consideration of containing P references (P_6 and P_9) for the determination of number of references for B_7 and B_8

Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US04/027432

International filing date: 24 August 2004 (24.08.2004)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/497,814
Filing date: 26 August 2003 (26.08.2003)

Date of receipt at the International Bureau: 08 October 2004 (08.10.2004)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)